

**UNITED STATES PATENT & TRADEMARK OFFICE**

Re: Application of: **Marian TRINKEL et al.**  
Serial No.: To Be Assigned  
Filed: Herewith  
For: **METHOD AND SYSTEM FOR CREATING SPEECH  
VOCABULARIES IN AN AUTOMATED MANNER**

**LETTER RE: PRIORITY**

Commissioner for Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

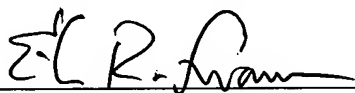
March 4, 2004

Sir:

Applicant hereby claims priority of German Application Serial No. 103 11 581.1, filed 10 March 2003.

Respectfully submitted,

DAVIDSON, DAVIDSON & KAPPEL, LLC

By   
Erik R. Swanson  
Reg. No. 40,833

Davidson, Davidson & Kappel, LLC  
485 Seventh Avenue, 14<sup>th</sup> Floor  
New York, New York 10018  
(212) 736-1940



## Prioritätsbescheinigung über die Einreichung einer Patentanmeldung

**Aktenzeichen:** 103 11 581.1

**Anmeldetag:** 10. März 2003

**Anmelder/Inhaber:** Deutsche Telekom AG, 53113 Bonn/DE

**Bezeichnung:** Verfahren und System zum automatisierten Erstellen von Sprachwortschätzen

**IPC:** G 10 L 15/06

Die angehefteten Stücke sind eine richtige und genaue Wiedergabe der ursprünglichen Unterlagen dieser Patentanmeldung.

München, den 13. Februar 2004  
**Deutsches Patent- und Markenamt**  
Der Präsident  
Im Auftrag

## **Verfahren und System zum automatisierten Erstellen von Sprachwortschätzen**

5 Die Erfindung betrifft ein Verfahren zum Aufbau und/oder zur Erweiterung einer Wortschatzdatenbank eines Spracherkennungssystems durch akustisches Trainieren des Spracherkennungssystems. Die Erfindung betrifft weiterhin ein Spracherkennungssystem mit einer Wortschatzdatenbank.

10 Spracherkennungssysteme sind allgemein bekannt und werden mittlerweile in verschiedenen Anwendungsgebieten eingesetzt. Beispielsweise kann ein Spracherkennungssystem eingesetzt werden, um in Abkehr von einer manuellen Bedienung eine Datenverarbeitungsanlage oder eine jegliche Maschine mittels Sprachbefehlen zu bedienen.

15

Ebenso gibt es Anwendungen in Form von sogenannten Diktierprogrammen, bei dem ein Spracherkennungssystem die von einem Benutzer in ein Mikrofon gesprochenen Worte analysiert, erkennt und in Textdaten umsetzt, so dass hierdurch ein direktes Diktieren von Text in eine Textverarbeitung eines

20 Rechnersystems ermöglicht wird.

Die Grundlage eines jeglichen solchen Spracherkennungssystems bildet eine Wortschatzdatenbank, die für einen Vergleich des von einem Benutzer gesprochenen Wortes mit dem gespeicherten Wortschatz herangezogen wird, um mit einer hohen Genauigkeit feststellen zu können, welches Wort ein Benutzer gesprochen hat und dementsprechend im Text umzusetzen ist.

25

Eine derartige Wortschatzdatenbank enthält nicht die Worte im eigentlichen Sinn, sondern Daten/Parameter, die aus gesprochenen Worten ermittelt wurden und grundsätzlich von der Art des Erkennungsalgorithmus abhängig sind, der einer Spracherkennung zugrunde gelegt wird.

30

So ist es bekannt, verschiedene Methoden der Spracherkennung einzusetzen, die beispielsweise häufig auf den sogenannten Hidden-Markov-Modellen oder dem „dynamic pattern matching“ bzw. „dynamic time warping“ beruhen, bei dem ein untersuchtes Wort mit Referenzwörtern verglichen wird, die im Wortschatz gespeichert sind.

Häufig haben die verschiedenen Möglichkeiten der Spracherkennung gemeinsam, dass ein erhaltenes Sprachsignal einer akustischen Vorverarbeitung unterzogen wird, bei der die Wörter in Phoneme unterteilt werden, d.h. in sprachlich kleinste Einheiten. Hierzu wird eine funktionelle Analyse der verschiedenen möglichen Laute einer Sprache vorgenommen.

Es können z.B. in einem ersten Schritt einer Spracherkennung Kurzzeitspektren eines akustischen Signals aufgenommen werden, die in einer Analyse direkt oder nach einer Datenverarbeitung als Muster dienen für einen Vergleich mit in einer Wortschatzdatenbank abgespeicherten Referenzmustern.

Unabhängig von der Art des Algorithmus bedarf es dementsprechend immer einer Wortschatzdatenbank bzw. deren Parameter mit einem für den verwendeten Algorithmus wortschatztypischen Aufbau, die zur Erkennung von gesprochenen Worten verwendet wird. Hierbei wird in Verbindung mit Spracherkennungsprogrammen oder -systemen üblicherweise eine Standardwortschatzdatenbank mitgeliefert, mit der bereits eine hohe Erkennungsquote der von einem Benutzer gesprochenen Worte möglich ist.

Häufig besteht jedoch noch die Notwendigkeit eine Wortschatzdatenbank für ein neues Sprachfeld zu erweitern, insbesondere dann, wenn Fachworte verwendet werden, die bislang in der Wortschatzdatenbank nicht zur Verfügung standen. Zur Aufnahme derartiger Fachworte bzw. allgemein von neu zu erlernenden Worten ist es üblicherweise vorgesehen ein Spracherkennungssystem akustisch zu trainieren, was bedeutet, dass dem Spracherkennungssystem die neu zu lernenden Worte vorgesprochen werden. Durch Aufnahme dieser neu

vorgesprochenen Worte in die Wortschatzdatenbank wird dementsprechend die Wortschatzdatenbank kontinuierlich vergrößert, so dass das Spracherkennungssystem einen neuen Wortschatz erlernen kann.

- 5 Im Stand der Technik ist es bekannt derartige Wortschatzdatenbanken üblicherweise durch hohen personellen Aufwand aufzubauen bzw. zu erarbeiten. Hierfür werden die neu aufzunehmenden Worte zusammengestellt, bearbeitet und in mühevoller personeller Arbeit durch Menschen z.B. in eine akustische Datenbank aufgesprochen, mit der dann ein Spracherkennungssystem auf die  
10 bekannte Art und Weise akustisch trainiert wird.

- Hierbei wird unter einem akustischen Trainieren nicht nur verstanden, dass neu zu lernende Worte zunächst in akustische Schallwellen gewandelt werden und über einen Mikrofoneingang einem Spracherkennungssystem zur Verfügung gestellt  
15 werden. Grundsätzlich kann beim akustischen Trainieren eines Spracherkennungssystems eine Schallwandlung unterbleiben und dem Spracherkennungssystem sofortig die akustischen Daten elektronisch zur Verfügung gestellt werden.

- 20 Dies ist z.B. der Fall, wenn eine Tonbandaufzeichnung direkt in den Mikrofoneingang eines Spracherkennungssystems elektronisch eingespielt wird, ohne vorher eine Schallwandlung vorzunehmen. Auch ein derartiges Trainieren eines Spracherkennungssystems wird im Sinne dieser Erfindung als akustisches Trainieren verstanden, da das Training auf akustischen, wenn auch nur  
25 elektronisch vorliegenden Signalen beruht.

- Das Aufbauen und Erweitern einer Wortschatzdatenbank und deren Parameter, wie im Stand der Technik bekannt, stellt dementsprechend einen großen manuellen und personellen Aufwand dar, so dass derartige Datenbanken nur unter  
30 Einsatz großer Kosten erstellt, erarbeitet und erweitert werden können.

Aufgabe der Erfindung ist es ein Verfahren und ein System zum Aufbau und/oder zur Erweiterung einer Wortschatzdatenbank eines Spracherkennungssystems bereitzustellen, mit dem auf kostengünstige Art und Weise ohne personellen Einsatz eine Wortschatzdatenbank aufgebaut oder eine bestehende erweitert werden kann.

Aufgrund des hohen personellen Aufwandes durch viele verschiedene Menschen ergeben sich im Stand der Technik auch Probleme im Trainingsprozeß des Spracherkennungssystems, da jeder Mensch ein anderes Stimmbild hat, welches mit dem der später bedienenden Person nicht übereinstimmt.

Diese Aufgabe wird gemäß der Erfindung dadurch gelöst, dass das Spracherkennungssystem durch ein computergestütztes Audiomodul trainiert wird.

Der Grundgedanke der Erfindung liegt darin, ein Spracherkennungssystem statt durch eine Person zu trainieren bzw. die Wortschatzdatenbank durch Personen zu erstellen / erweitern, die neu zu erlernenden Worte automatisiert vorzusprechen.

Gemäß der Erfindung ist es vorgesehen, dass dieses Vorsprechen neu zu erlernender Worte durch ein computergestütztes Audiomodul erfolgt.

Dementsprechend kann hier der personelle Aufwand minimiert werden, so dass die Wortschatzdatenbanken mittels des erfindungsgemäßen Verfahrens äußerst kostengünstig und standardisiert erstellt werden können.

Gemäß der Erfindung ist es bevorzugt vorgesehen, dass dem Audiomodul Wortschatzinformationen zugeführt werden, die das Audiomodul dem Spracherkennungssystem automatisch zur Erweiterung der Wortschatzdatenbank vorspricht. Wie oben erwähnt, setzt dieses Vorsprechen nicht notwendigerweise eine Wandlung der Wortschatzinformationen mittels eines Lautsprechersystems in Schall voraus, der dann anschließend mit einem Mikrofonsystem wiederum in ein elektrisches Signal gewandelt wird, sondern es besteht auch hier die Möglichkeit

die Schallwandlung zu umgehen und sofortig das akustische elektrische Signal dem Spracherkennungssystem zur Verfügung zu stellen.

5 Besonders bevorzugt wird bei dem erfindungsgemäßen Verfahren das Audiomodul die Wortschatzinformationen von einer Sprachdatenbank und/oder über ein Telekommunikationsnetzwerk erhalten. Gerade bei einer Lieferung der Wortschatzinformationen über ein Telekommunikationsnetzwerk besteht beispielsweise die Möglichkeit im sogenannten streaming modus die Daten zur Verfügung zu stellen. Dies kann beispielsweise über das Internet erfolgen, wenn  
10 z.B. Radiosendungen über das Internet empfangen werden. So besteht z.B. die Möglichkeit über das Internet das in einer Radiosendung verwendete Fachvokabular zu einem bestimmten Thema einem Spracherkennungssystem automatisch beizubringen, indem diese Streamingdaten dem Audiomodul zur Verfügung gestellt werden, welches sodann automatisch diese Sprachdaten dem  
15 Spracherkennungssystem vorspricht.

In einer weiteren bevorzugten Ausführung des erfindungsgemäßen Verfahrens kann es vorgesehen sein, dass die genannte Sprachdatenbank durch eine automatisierte Sprachsynthese von Textinformationen in einer  
20 Sprachsyntheseeinheit erzeugt wird. Hierbei können die Textinformationen z.B. einer Textdatenbank entnommen werden. Es besteht somit die Möglichkeit auf beliebige bestehende Textdatenbanken Rückgriff zu nehmen und die darin gespeicherten Textdaten durch eine Sprachsyntheseeinheit in Sprachinformationen umzuwandeln, die sodann in eine Sprachdatenbank  
25 eingeschrieben werden, welche ihrerseits dann wiederum dem Spracherkennungssystem zum Training zur Verfügung gestellt wird, wofür die in der Sprachdatenbank gespeicherten Sprachdaten z.B. über das Audiomodul dem Spracherkennungssystem vorgesprochen werden.

30 In einer besonders bevorzugten Ausführung kann das Audiomodul eines Spracherkennungssystems selbst eine derartige Sprachsyntheseeinheit umfassen, so dass Textinformationen, insbesondere aus einer Textdatenbank, direkt von dem

Spracherkennungssystem in Sprachinformationen gewandelt werden können, um anhand dieser Informationen das Training und somit die Erweiterung der Wortschatzdatenbank durchzuführen.

- 5 Die künstliche Sprachsynthese hat hier den Vorteil, dass das Spracherkennungssystem die Wortschatzinformationen immer mit einer „genormten“ Stimme vorgesprochen erhält, so dass weniger Probleme beim akustischen Trainieren vorkommen. Hierbei kann es vorgesehen sein, dass der Sprachsyntheseeinheit bestimmte gewünschte Sprach-Parameter bzw. Stimm-  
10 Parameter vorgegeben werden, z.B. hinsichtlich Geschlecht, der künstlichen Stimme, Alter, Körperbau, Dialekt, etc., um eine möglichst nahe Anpassung an den späteren tatsächlichen Benutzer des Spracherkennungssystems zu erreichen.

- Visuelle textliche Informationen können dem System automatisiert vorgegeben  
15 werden, z.B. durch ein Einscannen von Textdarstellungen.

- Neben der Möglichkeit vorhandene Textdatenbanken einzusetzen, kann das erfindungsgemäße Verfahren auch derart durchgeführt werden, dass die Textinformationen der Sprachsyntheseeinheit von einer automatisch erzeugten  
20 Textdatenbank zugeführt werden.

- Eine derartig automatisch erzeugte Textdatenbank kann für einen spezifischen Fall automatisch generiert werden, beispielsweise wenn Vokabular einer bestimmten Fachrichtung gezielt dem Spracherkennungssystem beigebracht werden soll.  
25 Hierfür kann es gemäß dem erfindungsgemäßen Verfahren bevorzugt vorgesehen sein, dass in der Textdatenbank automatisch die Textdaten zu wenigstens einer Textdatenquelle gespeichert werden, die über wenigstens eine Suchmaschine in einem internen oder externen Telekommunikationsnetzwerk, insbesondere dem Internet, zu wenigstens einem vorgegebenen Suchbegriff ermittelt werden.

30

Bekannt ist es, dass z.B. im Internet als einem möglichen externen Kommunikationsnetzwerk durch Eingabe eines gewünschten Suchbegriffes in



einer Suchmaschine eine Vielzahl sogenannter Links aufgefunden werden, unter denen Textdaten zu finden sind, die in engem Zusammenhang mit dem eingegebenen Suchbegriff stehen. Insofern können auf diese Art und Weise sehr schnell und vor allen Dingen kostengünstig erhebliche, z.B. statistisch relevante

5 Mengen von Textdaten aufgefunden werden, die thematisch mit dem Suchbegriff in Verbindung stehen und im Rahmen des erfindungsgemäßen Verfahrens dem Spracherkennungssystem zum Training zur Verfügung gestellt werden.

Hierfür kann es vorgesehen sein, dass durch eine Datenverarbeitungsanlage, gegebenenfalls durch das Spracherkennungssystem selbst, automatisch die

10 Textdaten von den ermittelten Textdatenquellen, also im Internet z.B. unter den verlinkten Adressen, gelesen und in der Textdatenbank gespeichert werden. So wird auf einfache und schnelle Art und Weise eine sehr große Textdatenbank aufgebaut, deren Inhalt mit dem Suchbegriff korreliert.

15 Da diese Textdaten gegebenenfalls auch solche Daten umfassen, die keinen Beitrag zur Wortschatzdatenbank liefern sollen, wie beispielsweise übliche Füllwörter oder Standardvokabular, kann es vorgesehen sein, dass die Textdaten in der Textdatenbank vor einer Sprachsynthese analysiert und bearbeitet werden.

20 Neben dem Entfernen von Füllwörtern kann es auch vorgesehen sein z.B. Mehrfachnennungen aus der Textdatenbank zu streichen und auch Informationen hinsichtlich einer Häufigkeitsverteilung bestimmter Worte zu erstellen, wobei diese Informationen ebenfalls in den Trainingsprozess des Spracherkennungssystems mit einfließen können, ebenso wie Informationen über Wahrscheinlichkeiten, mit

25 denen bestimmte Textdaten untereinander in Verbindung stehen.

So ist es beispielsweise bekannt bei einer Spracherkennung auch eine sogenannte Kontextprüfung durchzuführen, bei der ermittelt wird, mit welcher Wahrscheinlichkeit auf ein ermitteltes Wort ein anderes Wort folgt, um aus

30 mehreren möglichen Varianten eine sinnvolle Variante herauszusuchen. Dies wird beispielsweise durchgeführt, um Probleme mit Homophonen zu vermeiden, d.h. mit Worten, die gleich klingen, aber unterschiedliche Bedeutung haben.

Derartige Informationen z.B. über Kontextwahrscheinlichkeiten oder auch andere weitere Informationen jeglicher Art können gemäß der Erfindung vor der Durchführung einer Sprachsynthese aus den ermittelten Textdaten gewonnen und  
5 ergänzend einem Spracherkennungssystem zur Verfügung gestellt werden.

Gemäß der Erfindung umfasst dementsprechend im einfachsten Fall ein Spracherkennungssystem eine Wortschatzdatenbank sowie eine Sprachsyntheseeinheit, der zum Aufbau und/oder zur Erweiterung der  
10 Wortschatzdatenbank durch ein akustisches Vorsprechen Textdaten aus einer Textdatenbank zuführbar sind, wobei diese Textdatenbank erfindungsgemäß durch eine automatische Suche nach Textdaten zu einem vorgegebenen Suchbegriff in einem Telekommunikationsnetzwerk generiert ist.

15 Ein Ausführungsbeispiel der Erfindung ist in den nachfolgenden Zeichnungen näher erläutert. Es zeigen:

Figur 1 ein Spracherkennungssystem mit einem Anschluss an das Internet;

20 Figur 2 eine genauere schematische Darstellung eines Spracherkennungssystems

Die Figur 1 zeigt ein Spracherkennungssystem 1, welches Zugriff auf eine Wortschatzdatenbank 2 hat und von einem Benutzer 3 bedient wird. Ein derartiges  
25 System kann z.B. durch einen Heim-PC mit einem Diktierprogramm gebildet sein.

Neben der hier nicht weiter erläuterten Möglichkeit zur Spracherkennung z.B. im Rahmen einer Diktierfunktion innerhalb eines Textverarbeitungsprogramms ist gemäß der Erfindung das Spracherkennungssystem 1 über entsprechende  
30 Telekommunikationsleitungen mit dem Internet 4 verbunden.

Möchte nun ein Benutzer 3 den Sprachwortschatz in der Wortschatzdatenbank 2 z.B. durch spezifisches Fachvokabular erweitern, so kann der Benutzer 3 z.B. über ein Computerterminal dem Spracherkennungssystem einen Suchbegriff vorgeben, der in dem entsprechenden, neu zu erlernenden Fachbereich typisch ist. Über das erfindungsgemäße Spracherkennungssystem 1 wird mittels des Internets 4 z.B. eine erste Suchmaschine 5 angesprochen, in die der Suchbegriff eingetragen wird, woraufhin die Suchmaschine 5 im Internet bzw. in einer ihr zugeordneten Datenbank 6 nach Textdaten bzw. Hyperextdaten sucht, in denen der Suchbegriff vorkommt, woraufhin diese Textdaten ebenfalls wiederum über das Internet dem Spracherkennungssystem zur Verfügung gestellt werden.

Hier kann es auch vorgesehen sein, dass das Spracherkennungssystem 1 über das Internet zunächst eine zentrale Suchmaschine 7 mit der Suche nach dem gewünschten Begriff beauftragt, die ihrerseits auf mehrere Datenbanken 8 und 9 Zugriff hat und darüber hinaus noch weitere dezentrale Suchmaschinen 10 und 11 beauftragt, die ihrerseits in entsprechend zugeordneten Datenbanken nach dem Suchbegriff recherchieren. So kann also auch durch das Spracherkennungssystem eine sogenannte Meta-Suchmaschine beauftragt werden, die die Suche auf weitere Suchmaschinen unterverteilt.

Die insgesamt erhaltenen Textdatenmengen können dezentral oder auch zentral im Spracherkennungssystem gesammelt werden und gegebenenfalls nach einer Vorverarbeitung zum Trainieren des Spracherkennungssystems mittels einer Sprachsyntheseinheit herangezogen werden. Dieses Vorgehen ist in der Figur 2 näher erläutert.

Gemäß der Figur 2 kann ein Benutzer 3 über ein Computersystem 12 eine Suchbegriffsanfrage starten z.B. über eine Telekommunikationsverbindung in das Internet 4 an eine oder mehrere Suchmaschinen 5, die z.B. Datenbanken 6 Zugriff haben.

Die ermittelten Textquellen, im Internet als Links bezeichnet, werden gemäß dem erfindungsgemäßen Verfahren z.B. bevorzugt durch das Computersystem 12 automatisch aufgerufen, so dass die darin enthaltenen Textdaten gesammelt und in eine Textdatenbank 13 übermittelt werden können, wo diese Textdaten gesammelt und gegebenenfalls überarbeitet werden, z.B. in dem Sinne, dass Füllworte gestrichen, Mehrfachnennungen eliminiert und gegebenenfalls Kontextzusammenhänge festgestellt werden.

Die gesammelten, in der Textdatenbank 13 vorgehaltenen Textdaten können sodann einer Sprachsyntheseinheit 14 zugeführt werden, wodurch die Textdaten in Sprachinformationen konvertiert werden und in der Datenbank 2 abgelegt werden.

Nach dieser Sprachkonversion erfolgt die eigentliche Lernphase, d.h. dem Spracherkennungssystem 1 werden intern gegebenenfalls ohne eine Schallwandlung nur auf elektronischem Wege die Sprachdaten aus der Datenbank 2 vorgesprochen, wobei eine interne Datenbank des Spracherkennungssystem 1 erweitert wird.

Die einzelnen Elemente 1, 12, 13, 14 und 2 können auch zu einem Modul 15 zusammengefasst werden.

Das erfindungsgemäße Verfahren bietet dementsprechend zusammenfassend eine sehr kostengünstige Möglichkeit eine bestehende Wortschatzdatenbank eines Spracherkennungssystems bzw. eine neu aufzubauende Wortschatzdatenbank zu generieren, in dem automatisch auf eine Fülle von Textdaten der entsprechenden Datenbanken Rückgriff genommen wird, wobei diese Textdaten bevorzugt durch eine Sprachsyntheseinheit dem Spracherkennungssystem vorgesprochen werden, um den Lernprozess auszuführen.

### Patentansprüche

5

1. Verfahren zum Aufbau und/oder zur Erweiterung einer Wortschatzdatenbank (2) eines Spracherkennungssystems (1) durch akustisches Trainieren des Spracherkennungssystems (1), **dadurch gekennzeichnet, dass** das Spracherkennungssystem (1) durch ein computergestütztes Audiomodul trainiert wird.

10

2. Verfahren nach Anspruch 1, **dadurch gekennzeichnet, dass** dem Audiomodul Wortschatz-Informationen zugeführt werden, die das Audiomodul dem Spracherkennungssystem (1) automatisiert zur Erweiterung der Wortschatzdatenbank (2) vorspricht.

15

3. Verfahren nach einem der vorherigen Ansprüche, **dadurch gekennzeichnet, dass** das Audiomodul die Wortschatz-Informationen von einer Sprachdatenbank und/oder über ein Telekommunikationsnetzwerk (4), insbesondere im Streaming-Modus, erhält.

20

4. Verfahren nach Anspruch 3, **dadurch gekennzeichnet, dass** die Sprachdatenbank durch automatisierte Sprachsynthese von Textinformationen in einer Sprachsynthese-Einheit (14), insbesondere aus einer Textdatenbank (13), erzeugt wird.

25

5. Verfahren nach einem der vorherigen Ansprüche, **dadurch gekennzeichnet, dass** das Audiomodul eine Sprachsynthese-Einheit (14) umfasst, die Textinformationen, insbesondere aus einer Textdatenbank (13), in Sprachinformationen wandelt.
6. Verfahren nach Anspruch 5, **dadurch gekennzeichnet, dass** die Textinformationen der Sprachsynthese-Einheit (14) von einer automatisch erzeugten Textdatenbank (13) zugeführt werden.

7. Verfahren nach Anspruch 6, **dadurch gekennzeichnet, dass** in der Textdatenbank (13) automatisch die Textdaten zu wenigstens einer Textdatenquelle gespeichert werden, die über wenigstens eine Suchmaschine (5,7,10,11) in einem internen oder externen Telekommunikationsnetzwerk (4), insbesondere dem Internet, zu wenigstens einem vorgegebenen Suchbegriff ermittelt werden.
8. Verfahren nach Anspruch 7, **dadurch gekennzeichnet, dass** durch eine Datenverarbeitungsanlage automatisch die Textdaten von den ermittelten Textdatenquellen gelesen und in der Textdatenbank (13) gespeichert werden.
9. Verfahren nach einem der vorherigen Ansprüche, **dadurch gekennzeichnet, dass** die Textdaten in der Textdatenbank (13) vor einer Sprachsynthese (14) analysiert und bearbeitet werden.
10. Spracherkennungssystem mit einer Wortschatzdatenbank, **gekennzeichnet durch** eine Sprachsynthese-Einheit (14), der zum Aufbau und/oder zur Erweiterung der Wortschatzdatenbank (2) durch akustisches Vorsprechen Textdaten aus einer Textdatenbank (13) zuführbar sind, die durch eine automatische Suche nach Textdaten zu einem vorgegebenen Suchbegriff in einem Telekommunikationsnetzwerk (4) generiert ist.

### Zusammenfassung

- 5 Die Erfindung betrifft ein Verfahren zum Aufbau und/oder zur Erweiterung einer Wortschatzdatenbank (2) eines Spracherkennungssystems (1) durch akustisches Trainieren des Spracherkennungssystems (1), bei dem das Spracherkennungssystem (1) durch ein computergestütztes Audiomodul trainiert wird. Die Erfindung betrifft weiterhin ein Spracherkennungssystem mit einer
- 10 Wortschatzdatenbank, und einer Sprachsynthese-Einheit (14), der zum Aufbau und/oder zur Erweiterung der Wortschatzdatenbank (2) durch akustisches Vorsprechen Textdaten aus einer Textdatenbank (13) zuführbar sind, die durch eine automatische Suche nach Textdaten zu einem vorgegebenen Suchbegriff in einem Telekommunikationsnetzwerk (4) generiert ist.

15

Figur 1

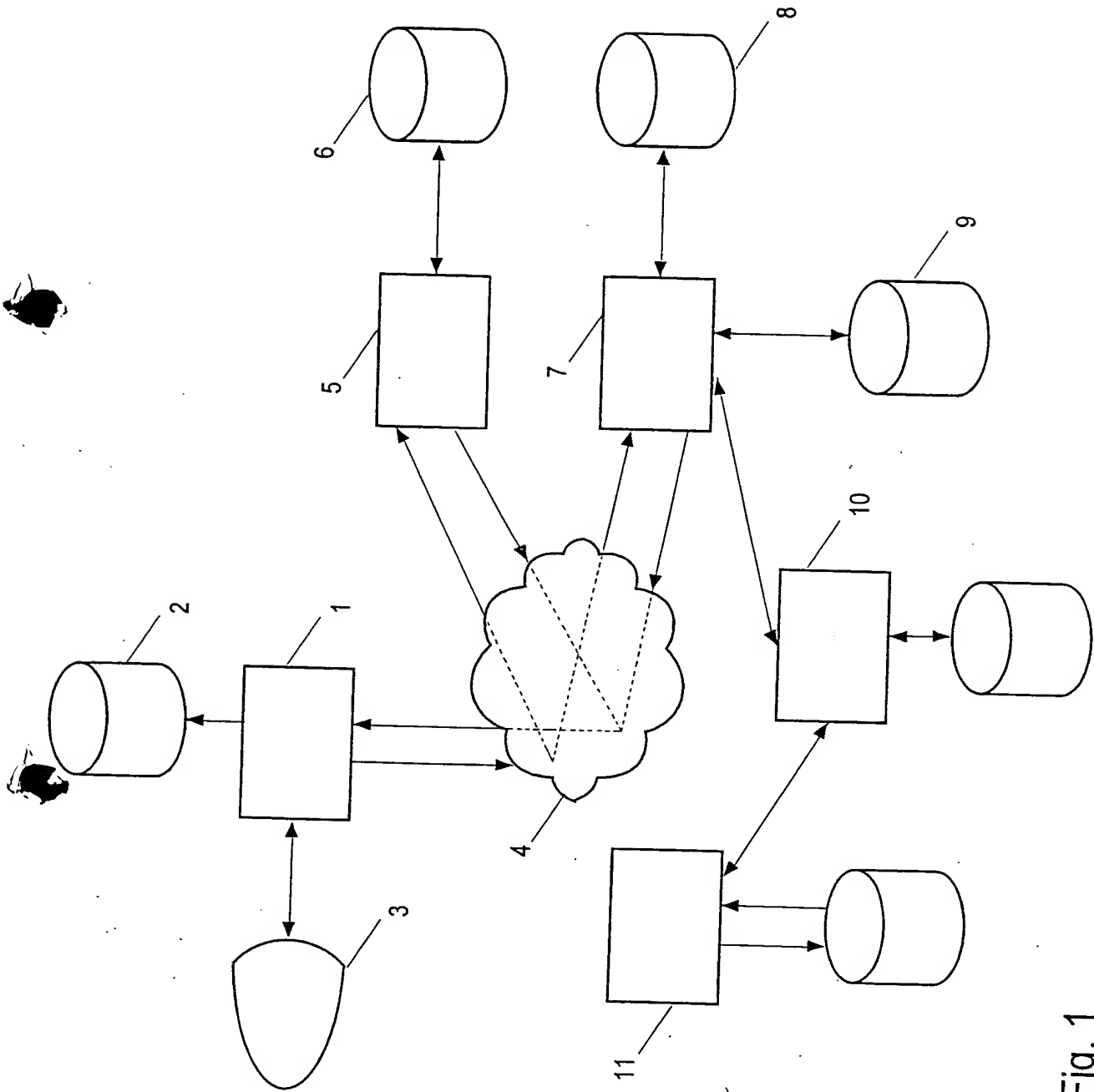


Fig. 1



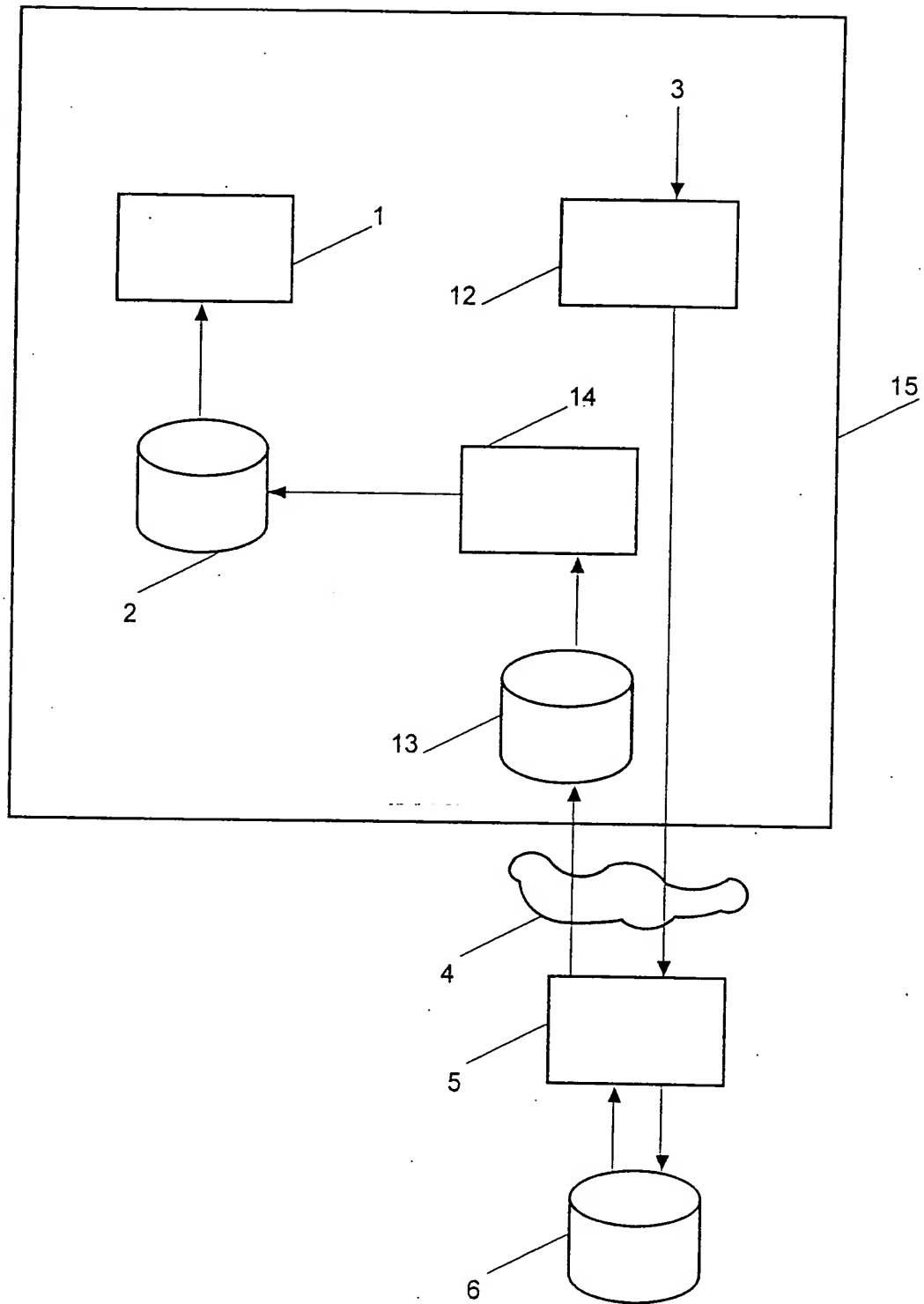


Fig. 2